

Gradedness and Consistency in Grammaticality Judgments

Aria Adli

1 The importance of graded grammaticality judgments: a case study of *que* → *qui* in French

The methodological issue of the unreliability of certain introspective data circulating in the syntactic literature has already been mentioned by several authors (e.g. Schütze 1996; Adli 2004). One particularly problematic phenomenon is that questionable judgments are sometimes quoted in theoretical studies without prior critical empirical verification, contributing to the formation of “myths” in the literature. One case is the *que* → *qui* ‘rule’ in French. This rule, which has been introduced into the literature solely on the basis of uncontrolled introspective data, is not confirmed by an experimental study in which a controlled process of data collection is applied to a whole sample of test subjects and which makes use of a graded concept of grammaticality.

The *que* → *qui* rule essentially states that an ECP violation can be avoided in French if *qui* is used instead of the usual complementizer *que* in sentences where a *wh*-phrase has been extracted from the subject position (see Perlmutter 1971; Kayne 1977). This rule rests on the empirical ‘premise’ that there should be a clear difference in grammaticality between (2a) and (2b) (all four sentences are taken from Hulk and Pollock 2001).

- (1) a. *Quel livre crois-tu que les filles vont acheter.*
which book think-you COMP_{que} the girls will buy
b. **Quel livre crois-tu qui les filles vont acheter.*
which book think-you COMP_{qui} the girls will buy
- (2) a. **Quelles filles crois-tu que vont acheter ce livre-là.*
which girls think-you COMP_{que} will buy that book-there
b. *Quelles filles crois-tu qui vont acheter ce livre-là.*
which girls think-you COMP_{qui} will buy that book-there

The *que* \rightarrow *qui* rule has been an often-used argument in syntactic theorizing.¹ The assumption is that this rule is a sort of loophole to avoid ungrammaticality, or in Pesetsky's words (1982: 308): "*Qui* does not occur freely as a complementizer, but only 'when needed' to avoid an NIC violation. [...] In other words, *qui* is a form of *que* which provides an 'escape hatch' from the effects of the NIC." Chomsky (1977) compares it with free deletion in COMP in English.

Rizzi (1990; 1997) supports his assumptions concerning the agreement process in the COMP system with this rule. He states that in cases of felicitous subject extraction in French the agreeing complementizer is not \emptyset , but the overt form *qui*. He assumes that an ECP violation is produced if the agreeing form does not occur and C is in what he considers as the unmarked form *que*. He further states that this rule is a morphological reflex of Spec-head-agreement between a trace and the head of COMP. Therefore Rizzi (1990: 56) assumes:

- (3) $qui = que + Agr$

Rizzi (1990) accounts for the ungrammaticality of the object extraction (1b) by assuming that Spec-head-agreement requires a C-adjacent position of the extracted element.

Furthermore, Rizzi (1990) assumes that the *que* \rightarrow *qui* rule only applies when agreement occurs between C^0 on the one hand and its specifier as well as its complements on the other hand. (Such a double agreement had already been described for Bavarian German by Bayer 1984 concerning sentences like *Wenn-st du kumm-st*).

The result would be as shown in (4): t' agrees with C^0 , t with I^0 and – due to the identity of t and t' – C^0 with the maximal projection of I^0 (by transitivity).

- (4) $[t' \ C^0 \ [t \ I^0 \ \dots]$

One aim of this paper was to test this assumption in an experimentally controlled process of data collection using a *graded* concept of grammaticality. Such a graded concept is assumed in Chomsky (1964), but it is already given up in Chomsky (1965) in favour of a distinction between grammaticality and acceptability. However, a rather pre-theoretic concept of gradedness persists in the syntactic literature, sometimes tacitly through the use of symbols like "?", "??", etc. Furthermore, some principles even make use of theoretical predictions in line with a graded concept (e.g. ECP vs. subjacency violation).

In order to measure graded grammaticality judgments, an instrument based on the principle of *graphic rating* (cf. Guilford 1954: 270; Taylor and Parker 1964) has been developed. Part of the design is an extensive instruction and training phase. Judgments are expressed by drawing a line on a bipolar scale (and not by marking one of several boxes with a cross). Within the limits of a person's differential capacity of judgment, a theoretically infinite number of gradations are therefore possible.

The test was presented in a A4 ring binder containing two horizontally turned A5 sheets (see diagram).

Figure 1

The upper sheet contained the reference sentence, the lower sheet the experimental sentence. The sentence, with the graphic rating scale under it, was printed in the middle of each sheet. After the subject had rated the experimental sentence on the lower sheet, he or she turned this page to go on with the next sentence. The upper sheet with the reference sentence was not turned and remained visible during the whole test. The judgments were given relative to the reference sentence judged in the beginning by the sub-

ject himself, within both endpoints (obviously well-formed and obviously ungrammatical) given by the design. It was, therefore, a bipolar, *anchored* rating scale with the characteristic that the subjects choose the anchor for themselves. The reference sentence consisted of a suboptimal, but not extremely ungrammatical, sentence. The dependent variable was the difference between the judgment of a particular sentence and the judgment of the reference sentence. The test started, after the presentation of written instructions, with an interactive instruction and training phase of about 10 to 15 minutes. During this phase, two main concepts were introduced in a 9-step procedure: isolated grammaticality and gradedness (cf. Adli 2004: 85-88 for details). A pre-test revealed the importance of such an additional training phase. Although not directly visible to the naked eye, the concept of grammaticality was often confounded with extra-grammatical factors (e.g., the plausibility of the situation described by the sentence). The understanding of the concept of isolated grammaticality is necessary to reduce interferences with semantic and pragmatic effects. Furthermore, subjects had to replace the common distinction between grammatical and ungrammatical, or "good" and "bad", sentences with a truly graded notion of grammaticality.

They were introduced to these two main concepts, among other things, by rating different training sentences and by explaining the reasons for their ratings to the experimenter, who could therefore adapt the instructions to the level of understanding of each subject. After instruction and training, the experimenter left the room.

Given that reliability can generally be improved by the use of several items, each syntactic structure was presented in 4 lexical variants.

Since the use of experimental methods in grammar research is recent, and not much experience exists yet, the evaluation of the instrument with regard to its reliability is important. A reliability analysis indicates the limits of an instrument concerning the precision of its measurements. Furthermore, the only three studies on the reliability of experimentally collected, graded grammaticality judgments I know of, namely Bard, Robertson and Sorace (1996: 61), Cowart (1997: 23) and Keller (2000: 215), rely on erroneous or improper calculations.²

Reliability is evaluated by Cronbach's α , which is a measure of internal consistency (see Cronbach 1951). It indicates the consistency between the different lexical variants of a sentence without taking into consideration mean differences between the variants. Indeed, the reliability of the measurements turned out to be sufficiently high (Cronbach's $\alpha = 0.85$).

78 French native speakers participated in the experiment. Validity was ensured by means of a special index (called *violation of trivial judgments*), reflecting the capability of the subject to give graded grammaticality judgments (cf. Adli 2004: 89-91). By means of this criterion, those subjects who were deemed unable to perform this task could be identified and excluded; the data of 65 subjects could be utilized for the subsequent statistical analyses.

Given that the measure of graded grammaticality does not reflect the categorical distinction between well-formed and ill-formed sentences, and given that such an information is still – for theory-internal reasons – important, grammatical as well as ungrammatical constructions were included in the test design in order to make available comparative scale points for the interpretation process: The experiment did not only cover subject-initial and object-initial interrogatives with long extraction over *que* and/or *qui*. The clearly felicitous constructions (5a) and (5b) with a PP-parenthetical “d’après vous” and the sentences (6a) and (6b) with the expression “croyez-vous” at the position after the *wh*-phrase were also included – some aspects of their syntax are discussed in section 3 (see Adli 2004 for full details).³

- (5) a. *Quel apache, d’après vous, méconnaît les obstacles de l’hiver?*
 which Apache according you ignores the difficulties of the
 winter
- b. *Quel animal, d’après vous, rôtiennent les esquimaux de l’ igloo?*
 which animal according you grill the Eskimos of the Igloo
- (6) a. (?)*Quel architecte, croyez-vous, conçoit les demeures du*
 président?
 which architect think you designs the residences of the
 president
- b. (?)*Quel argent, croyez-vous, investissent les organisateurs du*
 bal?
 which money think you invest the organisers of the
 ball
- (7) a. ??*Quel ingénieur, pensez-vous, qui conçoit la fusée de*
 l’Aérospatiale?
 which engineer think you qui_{COMP} designs the rocket of
 Aérospatiale
- b. **Quel idiot, pensez-vous, que perd les clefs de la maison?*
 which idiot think you que_{COMP} loses the keys of the house

- c. ?*Quel appel, pensez-vous, que* reçoivent les policiers
 du quartier?
 which call think you que_{COMP} receive the police officers
 of the district

The data was analysed with a two-way repeated measures ANOVA (variable A: “d’après vous” / “croyez-vous” / “pensez-vous qu-”; variable B: subject / object). I took into consideration not only information about the significance level, but also about the effect size of the differences (in terms of partial η^2 , cf. Cohen 1973; see also Keren & Lewis 1979: 119). The hypothesis was tested at $\alpha = 5\%$, which approximately allows for $\alpha = \beta$.⁴

In the following, only the relevant results concerning the *que* → *qui* issue will be given: In order to take into account the whole details of the results, a complete set of orthogonal simple effects was tested as regards the subject interrogatives (cf. Bortz 1999: 254), contrasting (i) (5a). vs. (6a), (ii) (7a) vs. (7b), as well as (iii) (5a) and (6a) vs. (7a) and (7b)

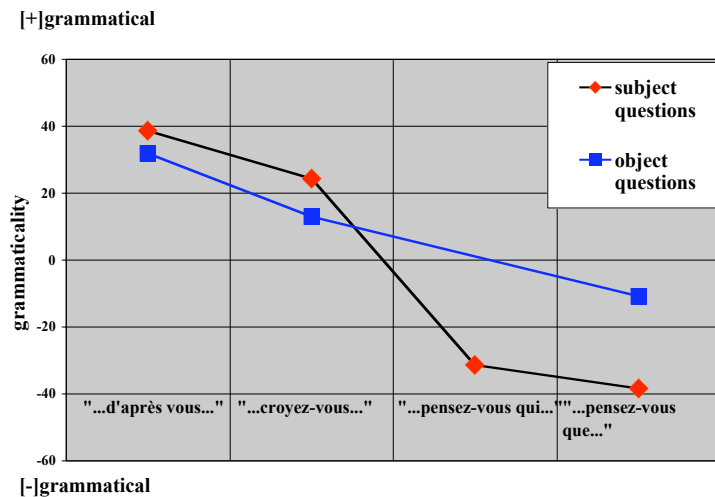


Figure 2

The results show a partial η^2 of 0.183 ($p < 0.000$) for contrast (i), a value of 0.149 ($p < 0.001$) for contrast (ii), but an amount as high as 0.875 ($p < 0.000$) for contrast (iii). It appears that the *qui*-form (7a) is anything but felicitous. Though there is a significant difference between the *qui*-form (7a) and the *que*-form (7b) (i.e., the ungrammaticality of the construction with *qui* is not as sharp as the ungrammaticality of the construction with *que*).

que), it is a matter of fine-grained differences *within* the range of ungrammatical constructions. The set of orthogonal simple effects shows that the different subject-initial constructions divide into two clearly separated groups, with an eye-catching decrease in grammaticality between them.

The form with *qui* thus cannot be considered as the licensed counterpart of the form with *que*. The *que* → *qui* rule emerges as a myth, and it must consequently be eliminated from the discussion. All the same, Pesetsky (1982: 308) notes that “for some French speakers” the use of *qui* does not make the sentence grammatical. However, he assumes that these persons are speakers of particular dialects (without specifying them) and he does not therefore cast doubt on the *que* → *qui* rule. As to the question why the ungrammaticality of the *qui*-form is less sharp, I suggest rather psycholinguistic factors to be responsible: the use of *qui* instead of *que* evokes the structure of subject relative clauses (i.e. the nominative *qui* has a sort of resumptive character), which alleviates the repair mechanisms. Concerning this, it is interesting to note that Perlmutter (1971), the first to raise the *que* → *qui* issue, analyzes *qui* in sentences with long subject extraction as a relative pronoun, as his gloss to (8) shows.

- (8) *Qui a -t-il dit qui s'est évanoui?*
 who did he say who fainted
 ‘Who did he say fainted?’

Another argument in favour of the assumption of alleviated repair mechanisms is, at least for declaratives, the relatively easy re-analysis of these constructions: The expression *qu’il a dit* in example (9), also taken from Perlmutter (1971: 102), can be omitted (along the lines of a parenthetical analysis). The remaining sentence (10) is a usual construction with a relative clause. In addition, the expression *qu’il a dit* itself in (9) is not well-formed (if at all, an expression with the PP-pronoun *dont* would be required), favouring a reanalysis of the whole sentence with deletion of this expression. A similar situation can be stated for (11), taken from Rizzi (1990: 56), to give another example from the relevant literature.

- (9) *la speakerine qu’il a dit qui s’est évanouie...*
 the spokeswoman that he has said who fainted...
- (10) *la speakerine qui s’est évanouie...*
 the spokeswoman who fainted...
- (11) *l’ homme que je crois qui viendra...*
 the man that I believe who come-FUT

2 Graded grammaticality and the measure of judgment consistency

It is not surprising that the measure of judgment consistency has been so far ignored in syntactic research, essentially because its calculation requires a metrical (i.e., graded) grammaticality scale. The procedure is similar to the reliability evaluation of the instrument described in the previous section. However, for this purpose, the reliability values are interpreted *separately* for each construction. The main assumption is that reliability differences between different syntactic structures, measured with the same instrument under the same conditions, do not represent a mere indicator with respect to the precision of the instrument, but constitute an interpretable measure in terms of grammar theory. The approach of measuring graded grammaticality judgments allows one not only to study the mean value for the judgments in a sample, but also to calculate the internal judgment consistency (one might also say “intra-individual judgment consistency”) and to compare these values for different syntactic structures. This measure has the advantage of complementing the information about the exact grammaticality value with the information on the difficulty of giving stable judgments, allowing a more complete evaluation of the grammatical quality of a structure.

I conducted reliability analyses using the average-measure intraclass correlation coefficient (ICC) of the absolute agreement type (cf. McGraw and Wong 1996). This value indicates the intra-individual degree of agreement between the judgments of the lexical variants for each construction.

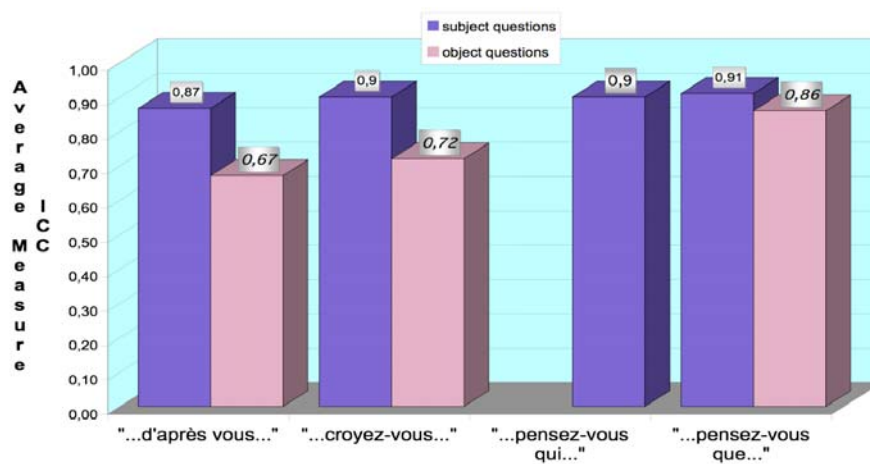


Figure 3

Taking into account differences in mean between the lexical variants, the ICC, derived from the analysis of variance, is a more severe (or conservative) measure than Cronbach's α (the specific form applied is the two-way model with random variables).

The results given in the figure show that (i) consistency of grammaticality judgments is not a stable factor but depends on the respective construction, (ii) in terms of our examples, both subject extractions, namely with *qui* and with *que*, have comparable consistency values, and (iii) in French it is more difficult to give consistent judgments to object interrogatives than to subject interrogatives.

Consistency in the judgment of the object interrogatives improves with increasing suboptimality, as a comparison of the last two figures reveals: Consistency is much higher for (7c) than for (5b) or (6b), i.e. there is an *interaction* between the degree of suboptimality and the sentence initial element. Hence, the analysis of judgment consistency provides another piece of empirical evidence in terms of the discussion about the syntax and the processing of subject- vs. object-initial interrogatives in French. Given this result, two further questions arise concerning (i) the general difference between subject-initial and object-initial interrogatives and (ii) the pronounced suboptimality of the long object extraction (7c). These issues show how grammaticality values, as well as consistency values, come into play in a discussion.

3 Further issue: extraction, parenthesis and analogy

3.1 The contrast between subject-initial and object-initial questions

The results of the judgment consistencies are in line with the results of the judgment values itself: Simple main effect tests of the variable B (subject vs. object) reveal significant for each of the three construction types:⁵

The difference between subject- and object-questions with the PP-parenthetical “d’après vous”, (5a) and (5b), is significant. Subject questions have a higher grammaticality value than object questions ($p < 0.034$; partial $\eta^2 = 0.068$).⁶

The difference between subject- and object-questions with “croyez-vous”, (6a) and (6b), is significant. Subject questions have a higher grammaticality value than object questions ($p < 0.000$; partial $\eta^2 = 0.271$).⁷

The difference between subject- and object-questions with long extraction, (7b) and (7c), is also significant. However, subject questions have in

this case a lower grammaticality value than object questions ($p < 0.000$; partial $\eta^2 = 0.574$).⁸

I assume psycholinguistic factors, having an impact on grammaticality judgments, to be responsible for this effect. Apart from a few exceptions (e.g. Farke 1994), a more difficult processing of object-initial sentences has often been claimed in the psycholinguistic literature (see Frazier and Flores d'Arcais 1989 for Dutch, de Vincenzi 1991 for Italian, Hemforth 1993 for German, cf. also Gorrell 2000). In line with Schütze (1996: 164) who claims that "any other factors that might make a sentence hard to parse" affect the judgment, I also assume that the more difficult processing of the French object-initial questions affects the judgments. The unambiguous interpretation of French subject-initial and object-initial interrogatives like (5a) through (7c), especially the correct interpretation of object-initial interrogatives, requires particular morphological, semantic and phonetic cues. Unlike German, French is not a case language.⁹

Other than the difference between (7b) and (7c), the difference in grammaticality between subject-initial and object-initial questions without long extraction, e.g. (5a) and (5b), is anything but trivial. So far, not much attention has been paid to differences in grammaticality between *licensed* constructions.

Concerning these results, the follow-up question arises as to whether marked forms, as long as they are not clearly suboptimal, generally have a lower judgment consistency than their unmarked counterparts. Future research might give an answer to this issue.

3.2 Long object extraction and analogy

We have stated so far that consistency in the judgment of the object interrogatives is lower compared to subject interrogatives. However, this only applies as long as the construction is not clearly suboptimal. The long object extraction (7c) has a much higher consistency than the two other object-initial constructions (5b) and (6b). The question arises why long object extractions are clearly suboptimal in French. Contrasting with the long subject extraction (7b) violating the ECP-condition, there is no obvious reason for explaining the low grammaticality value of long object extractions.

In order to address this issue, we first need to turn to the constructions (6a) and (6b). I will call these constructions with the expression "croyez-vous" right-adjacent to the *wh*-element, as suggested in Adli (2004), *VtoC-constructions*, contrasting with the long extraction cases like (7b) and (7c),

which I call *VImC-constructions*.¹⁰ We have already stated that VIoC-constructions are slightly suboptimal: The plot of the grammaticality values showed that they have a slightly lower degree of grammaticality than the sentences (5a) and (5b) with the PP-parenthetical “d’après vous”. In Adli (2004), I had already addressed the question as to whether French VIoC-constructions should be analyzed as simple matrix clauses with a parenthesis or as complex clauses with long extraction. I give a brief sketch of those results, however without entering into the details here: I assume sentences like (6a) and (6b) to be instances of simple clauses with a parenthesis. VIoC-expressions like “croyez-vous” exhibit certain properties characteristic for parenthetical constructions: (i) they can be omitted, (ii) they are restricted to the root position, (iii) they can appear in various positions in the sentence. (The same holds if the expression in question does show the canonical word order like “vous croyez” or if the object-initial question (6b) does not exhibit stylistic inversion). (6a) and (6b) being analyzed as parenthetical constructions, their slight suboptimality is assumed to be due to characteristics of the parenthesis (and not, for example, due to any movement operation). The comparison with the completely felicitous sentences (5a) and (5b) with the PP-parenthetical “d’après vous” suggests that the reason for the slight suboptimality of (6a) and (6b) resides in the fact that *sentencial* parentheses like “croyez-vous” are not permitted in French interrogatives. The sentencial property is related to the issue of the interpretive relation between the predicate and the object argument, the problematic point being the fact that the argument of VIoC-expressions like “croyez-vous” has to be specified by the host sentence (cf. Reis 1995; 1996 concerning German). We can observe that French declaratives with a sentencial parenthesis are completely felicitous, when they occur with an overtly realized object. Interestingly, their variants with interpretive integration (i.e. without an overtly realized object) show the same slight suboptimality effect.¹¹

- (12) a. *Cet écrivain, on le sait, était un bon-vivant.*
 this writer one it knows was a bon vivant
 ‘This writer, as is generally known, was a bon vivant.’
 b. (?)*Cet écrivain, on sait, était un bon-vivant.*
 this writer one knows was a bon vivant
- (13) a. *Cette maison, comme vous le savez, est très ancienne.*
 this house as you it know is very old
 ‘This house is, as you know, very old.’

- b. (?) *Cette maison, comme vous savez, est très ancienne.*
 this house as you know is very old

One could assume that in French the grammar generally selects the form with a sentential parenthesis without interpretive integration as the „better candidate“. However, sentential parentheses without interpretive integration are not possible in French interrogatives.

- (14) a. **Où, le penses-tu, habite-t-elle avec l' enfant depuis 1985?*
 where it thinks you lives she with the child since 1985
 b. **Où, tu le penses, habite-t-elle avec l' enfant depuis 1985?*
 where you it thinks lives she with the child since 1985

The slight suboptimality of VIoC-constructions like (6a) and (6b) is therefore due to the fact that the form with interpretive integration, which is actually required in French, is not available in interrogatives and that the slightly suboptimal, integrated variant has to be used.¹²

This being said, we can turn back to the initial question as to why the long object extraction (7c) shows a high degree of suboptimality. A first intuition consists in the assumption of some kind of relationship between the (slight) suboptimality of object-initial VIoC-constructions like (6b) and the (strong) suboptimality of object-initial VImC-constructions like (7c). However, according to the present analysis, (6b) is a parenthetical construction and (7c) an extraction construction. It does therefore not seem easy to establish a relationship between them.

Interestingly, Reis (2000a; 2000b) has shown on the basis of the characteristics of German *was...-w*-constructions like (15a) that properties of extraction constructions like (15b) and properties of parenthesis constructions like (15c) can co-occur. In other words, parenthesis and extraction constructions are two related types of construction affecting each other – in a rather unorthodox manner – leading to hybrid phenomena.

- (15) a. *Was glaubst du, was er kochen sollte?*
 what believe you what he cook should
 b. *Was glaubst du, dass er kochen sollte?*
 what believe you that he cook should
 c. *Was sollte er glaubst du kochen?*
 what should he believe you cook

Reis (2000a: 28) enumerates several properties of *was...-w*-constructions that are typical for extraction constructions, e.g. the fact that

the *was*-clause is always initial, that the related *wh*-clause must contain a *wh*-moved *wh*-phrase (and hence is not an *ob*-interrogative), that the *was*...*w*-construction can be embedded, that the *was*-clause may contain more complex verbs of saying, thinking or believing like *behaupten* (claim) or *argwöhnen* (suspect), etc. At the same time, she points out several properties typical for parenthesis constructions, e.g. the fact that only those predicates can appear as bridge verbs in *was*...*w*-constructions which can also appear in parenthetical *was*-sentences like (15c), that complex predicates involving *es* or full object NPs like *scheint es* (seems it) or *hat sie das Gefühl* (has she the feeling) are allowed, contrasting with extraction constructions. Therefore she claims that parenthesis and extraction constructions are related and that properties of the one can be transferred on the other by processes of *analogy*: „Since a convincing account of EC-IPC ‚blends‘ [EC = extraction construction, IPC = integrated parenthetical construction, A.A.] is quite hard to give [...] it might seem better after all to treat the rather slight EC-traits of EV2 constructions as mere analogical transgressions [...] of the basic IPC pattern which the formal and interpretive closeness of prefinite IPCs to ECs gives rise to“ (Reis 2000a: 27).¹³

Sternefeld (1998: 28) also takes into account the idea of interpretive closeness in his analysis of German *was*...*w*-constructions. He applies the concept of semantic parallelism referring to the compositional semantics of Dayal (1991) for (simple) Hindi *wh*-constructions. He assumes a relationship between German *was*...*w*-constructions and the semantically parallel colonconstructions. This assumption helps him to give at least a partial account for the ungrammaticality of (17a) and (17b) on the basis of the idea of analogy. Sternefeld (1998) tries to explain why (16) is grammatical, but not (17a) and (17b), although multiple questions are generally possible in German.

- (16) *Was glaubst du, wer gekommen ist?*
 what believe you who come is
- (17) a. **Was glaubt wer, wer gekommen ist?*
 what believes who who come is
 b. **Wer glaubt was, wer gekommen ist?*
 who believes what who come is
- (18) a. **Was glaubt wer: Wer ist gekommen?*
 b. **Wer glaubt was: Wer ist gekommen?*

Sternefeld (1998) essentially argues that if we cannot explain the ungrammaticality of the hypotactic constructions (17a) and (17b), we should

examine the semantically equivalent, paratactic constructions (18a) and (18b) more thoroughly. It appears, that yet the paratactic construction is ungrammatical. This observation still does not give a complete account for the asymmetry between (16) and (17a), (17b). However, we can assume that whatever is responsible for the ungrammaticality of (18a) and (18b), is also responsible for the ungrammaticality of (17a) and (17b), because (17a), (17b) and (18a), (18b) are semantically parallel.

The idea of analogy underlying Sternefeld's analysis and also developed by Reis in the scope of her analysis of German *was...w*-constructions – namely “analogy rather than *wh*-movement plays the major role in accounting for long *wh*-extraction constructions” (Reis 2000b: 403) – also offers a possible explanation for the suboptimality of long object extractions in French. We can assume that whatever is responsible for the suboptimality of French VIoC-constructions like (6a) and (6b) is also responsible for the suboptimality of long object extractions like (7c). In other words: Even though the suboptimality of VIoC-constructions is due to reasons *specific to parenthetical constructions* (such as the assumption of the inadmissibility of sentential parentheses with interpretive integration in French), they also affect the VImC-construction by virtue of the closeness of parenthetical constructions and extraction constructions.

However, one still needs to account for the fact that the suboptimality is more pronounced for the object-initial VImC-construction than for the object-initial VIoC-construction in French. The first must therefore be affected by an additional factor, causing a decrease in grammaticality not affecting the latter. I assume that in many languages the long object extraction shows a lower degree of grammaticality than corresponding VIoC-constructions. This fact has not yet received much attention in the literature – possibly because *gradedness* has been considered for a long time as an epiphenomenon, but maybe also because the long object extraction often serves as a *counterpart* to the clearly ungrammatical long subject extraction (leading to the effect that grey next to black seems whiter than grey next to white). Along these lines I also assume German long object extractions as in (19b) to have a lower degree of grammaticality than (19a).

- (19) a. *Welchen Anruf glaubst du erhielt der Anwalt meiner Frau?*
 which call believe you received the lawyer of my wife
 b. *Welchen Anruf glaubst du, dass der Anwalt meiner Frau erhielt?*
 which call believe you that the lawyer of my wife received

In sum, we have seen that the controlled measurement of a graded concept of grammaticality does, on the one hand, allow one to obtain a fairly

detailed picture of grammaticality contrasts. Syntactic discussions can thus be placed on a more solid empirical base. This reduces the risk of myth production as in the *que-qui* case, and it allows reliable assumptions on fine-grained differences as the contrast between subject-initial and object-initial questions or the different degrees of suboptimality between object-initial VIoC- and VImC-constructions show. On the other hand, the same measurements can be analyzed from another point of view, namely with respect to judgment consistency. Judgment consistency seems to correlate with the degree of grammaticality as well as with certain structural properties, e.g. the initial element. It is a new and complementary source of information, to be hence worthy considered in grammar research.

Notes

1. Concerning this Rizzi (1990: 56) writes: „A significant body of work has been devoted to the rule converting *que* into *qui* in French *wh*-constructions.“
2. To put it briefly, they confound cases and variables, i.e. they calculate a correlation for a sample of variables and not for a sample of persons. By calculating the mean of the judgments of a sentence for all persons they eliminate the variance within the sample. Rather, the test-retest-reliability should have been calculated for each sentence separately. If desired, the mean of the different reliabilities could then have been calculated (taking into account the Fisher-Z-transformation). In addition, Bard et al. (1996: 23) and Keller (2000) compared two independent samples. Rather, the test-retest-reliability is defined as the correlation of two repeated measurements with the same sample. Only these mistakes explain why Bard et al (1996) obtain $r = 0.89$, Keller (2000: 217) $r = 0.90$, and Cowart (1997) even a hardly realistic $r^2 = 0.97$. Note that in the present study reliability is not calculated using the test-retest-reliability but by Cronbach's α .
3. Anticipating the results of the judgment test, the degree of grammaticality of the sentences (5a) to (7c) is indicated by the symbols (?), ?, ?? and *, roughly meaning “slightly suboptimal”, “suboptimal”, “highly suboptimal” and “ungrammatical”. This categorization is a simplification of the more detailed, metrical grammaticality values shown in the line chart and therefore does not convey the richness and precision of information of the gradedness approach. However, these categories are not the result of a mere recoding of the metrical values, but rather an *interpretation* of the values in terms of a categorical concept of well-formedness.
4. I consider α and β equally important with this issue. In other words, the conclusion that the grammaticality of two constructions is identical (i.e. a non-significant result) and the conclusion that the grammaticality of certain con-

structions are different (i.e. a significant result) has the same practical impact for the purposes of grammar research and should come along with the same error probability.

5. The overall main effect B itself cannot be interpreted because of a hybrid interaction effect A x B ($p < 0.000$), i.e. both levels of variable B show the same decreasing trend whereas the three levels of variable A do not show the same trend (cf. also Bortz 1999: 289-291).
6. Pillai's $PS_{B|a1} = 0.068$; $F = 4.7$; $df = 1$; $df_{error} = 64$; partial $\eta^2 = 0.068$; $p < 0.034$
7. Pillai's $PS_{B|a2} = 0.271$; $F = 23.842$; $df = 1$; $df_{error} = 64$; partial $\eta^2 = 0.271$; $p < 0.000$
8. Pillai's $PS_{B|a3} = 0.574$; $F = 86.093$; $df = 1$; $df_{error} = 64$; partial $\eta^2 = 0.574$; $p < 0.000$
9. Different phonetic, morphological and semantic disambiguation cues were combined in the design of the material: (i) a lack of agreement concerning the number feature between the verb and the object, (ii) a morpho-phonetic realization of number, i.e. not only a readable but also an audible subject-object-distinction, for the *wh*-element (by means of *liaison*) and for the verb (3rd group of conjugation), and (iii) a semantically founded assignment of the subject and object function in terms of selection constraints.
10. VIoC is the abbreviation derived from the German expression for "verb-initial sentential expression without COMP", and VImC the abbreviation derived from the German expression for "verb-initial sentential expression with COMP". These expressions shall be descriptive and neutral terms, especially with regard to the theoretical issue as to whether VIoC-constructions like (6a) and (6b) are to be analyzed as instances of long extraction or of simple matrix clauses with a parenthesis.
11. French and German are complementary with respect to the condition of interpretive integration in declaratives. The forms with interpretive integration are preferred in German.

(A) ?? *Karl begann, wie er das gesagt hatte, zu schreiben.*
Charles began as he this said had to write

(A') *Karl begann, wie er gesagt hatte, zu schreiben.*
Charles began as he said had to write
12. An explanation for the unavailability of the form with interpretive integration in French interrogatives has yet to be found.
13. Reis (1995) calls German sentences like (B) EV2-constructions. She essentially states that they are not, as has been often assumed, instances of extraction but rather a particular form of parenthetical construction.

(B) *Was glaubst du sollte er kochen?*
what believe you should he cook

References

- Adli, Aria
 2004 *Grammatische Variation und Sozialstruktur*. (= Studia Grammatica 58). Berlin: Akademie Verlag.
- Bard, Ellen Gurman, Dan Robertson and Antonella Sorace
 1996 Magnitude Estimation of Linguistic Acceptability. *Language*, 72 (1): 32-68.
- Bayer, Josef
 1984 COMP in Bavarian. *Linguistic Review*, 3: 209-274.
- Bortz, Jürgen
 1999 *Statistik für Sozialwissenschaftler*. Springer, Berlin, New York.
- Chomsky, Noam
 1964 *Current Issues in Linguistic Theory*. Mouton, Den Haag.
 1965 *Aspects of the Theory of Syntax*. MIT Press, Cambridge (Mass.).
 1977 On wh-movement. In Culicover, Peter, Thomas Wasow and Adrian Akmajian (eds.). *Formal syntax*. pp. 71-132. Academic Press, New York.
- Cohen, Jacob
 1973 Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33: 107-112.
- Cowart, Wayne
 1997 *Experimental Syntax. Applying Objective Methods to Sentence Judgments*. London: Sage Publications.
- Cronbach, Lee J.
 1951 Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297-334.
- Dayal, Veneeta S.
 1991 *Wh-Dependencies in Hindi*. PhD thesis, Cornell University.
- De Vincenzi, Marica
 1991 *Syntactic Parsing Strategies in Italian*. Kluwer, Dordrecht.
- Farke, Hildegard and Sascha W. Felix
 1994 Subjekt-Objektasymmetrien in der Sprachverarbeitung - Spurensuche. In Felix, Sascha W., Christopher Habel and Gert Rickheit (eds.). *Kognitive Linguistik - Repräsentationen und Prozesse*. Westdeutscher Verlag, Opladen.
- Frazier, Lyn and Flores D'Arcais, Giovanni B.
 1989 Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, 28: 331-344.
- Gorrell, Paul
 2000 The subject-before-object preference in German clauses. In Hemforth, Barbara and Lars Konieczny (eds.). *German Sentence Processing*. pp. 25-63. Dordrecht: Kluwer.

- Guilford, Joy P.
1954 *Psychometric methods*. MacGraw-Hill, New York.
- Hemforth, Barbara
1993 *Kognitives Parsing: Repräsentationen und Verarbeitung sprachlichen Wissens*. Infix, Sankt Augustin.
- Hulk, Aafke and Jean-Yves Pollock
2001 Subject positions in Romance and the theory of Universal Grammar. In Hulk, Aafke and Jean-Yves Pollock (eds.). *Inversion in Romance*. pp. 3-20. Oxford University Press, Oxford.
- Kayne, Richard S.
1977 French relative que. In Luján, Marta and Fritz Hensey (eds.). *Current Studies in Romance Linguistics*. Washington D.C.: Georgetown University Press.
- Keller, Frank
2000 *Gradience in Grammar. Experimental and Computational Aspects of Degrees of Grammaticality*. PhD dissertation, University of Edinburgh.
- Keren, Gideon and Charles Lewis
1979 Partial omega squared for ANOVA designs. *Educational and Psychological Measurement*, 39: 119-128.
- Levelt, Willem J. M.
1974 *Formal Grammars in Linguistics and Psycholinguistics, Vol. III*. Mouton, Den Haag.
- McGraw, Kenneth O. and Wong, S.P.
1996 Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1): 30-46.
- Perlmutter, David
1971 *Deep and Surface Structure Constraints in Syntax*. Holt, Rinehart and Winston, New York.
- Pesetsky, David
1982 Complementizer-trace phenomena and the Nominative Island Condition. *The Linguistic Review*, 1: 297-344.
- Reis, Marga
1995 Wer glaubst du hat recht? On so-called extractions from verb-second clauses and verb-first parenthetical constructions in German. *Sprache und Pragmatik*, 36: 27-83.
1996 Extractions from verb-second clauses in German?. In Lutz, Uli and Jürgen Pafel (eds.). *On Extraction and Extraposition in German*. pp. 45-88. John Benjamins: Amsterdam.
2000a Wh-movement and integrated parenthetical constructions. In Zwart, C. Jan-Wouter and Werner Abraham (eds.). *Studies in Comparative Germanic Syntax. Proceedings from the 15th Workshop on Comparative Germanic Syntax*. John Benjamins: Amsterdam.

- 2000b On the parenthetical features of German was-w-constructions and how to account for them. In Lutz, Uli, Gereon Müller, and Arnim v. Stechow (eds.) *Wh-Scope Marking*. pp. 359-407. John Benjamins: Amsterdam.
- Rizzi, Luigi
 1990 *Relativized Minimality*. MIT Press, Cambridge (Mass.).
 1997 The fine structure of the left periphery. In Haegeman, Liliane (ed.). *Elements of Grammar*. pp. 281-337. Kluwer: Dordrecht.
- Schütze, Carson T.
 1996 *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. University of Chicago Press, Chicago.
- Sternefeld, Wolfgang
 1998 Grammatikalität und Sprachvermögen. *SfS-Report* 02-98. Universität Tübingen.
- Taylor, James B. and Howard A. Parker
 1964 Graphic ratings and attitude measurement: a comparison of research tactics. *Journal of Applied Psychology*, 48(1): 37-42.